Serving Files with the US CMS File Server

Jim Hanlon

US CMS Project Office

Fermilab

August 13, 1998

File Serving

To effectively serve files, it is necessary to understand both how files are served by the server and, perhaps more importantly, how the files will be received and used by the “customer” or client

Considerations include the protocol (ftp, http, ...), the file type (.html, .txt, .ps, .doc, .xls, .mpp, ...) and size, and the assumed network and application software (and version) available to the client

Protocols

- Many computer-to-computer communications utilize tcp/ip (transmission control protocol/ internet protocol) protocols (e.g., the “internet”)
- Numerous other communication protocols are transmitted using tcp/ip
  - http - hypertext transport protocol
  - ftp - file transfer protocol
  - pop3 - post office protocol
  - smtp - simple mail transport protocol
  - nntp - network news transport protocol
  - ntp - network time protocol
- The different protocols use different IP “ports”

Servers and Clients

It is important to distinguish between server and client functions

Clients:

- **http (WWW): Netscape Navigator or Microsoft Internet Explorer**
- **ftp: Exceed, QVT/FTP, ... (Windows); Fetch, Anarchie, ... (Mac)**
  - **Netscape Navigator and Microsoft Internet Explorer have limited ftp capabilities; Internet Explorer lacks upload capabilities**
- **pop3: Eudora, Outlook, Navigator, Explorer, ...**

Servers (daemons):

- **A daemon is a server application program which runs on the host computer, and serves a particular protocol (e.g., ftp or http) to the network**

IP Names and Addresses

- Each computer on the internet has a unique hardware (ethernet) address specified by the manufacturer, usually given in dotted hexadecimal format
  - For example, the hardware address of the US CMS server is 08.00.07.6E.2B.BE
- Computers are assigned unique IP (internet protocol) names and addresses
- An IP address is specified in dotted decimal form (e.g., 131.225.52.112); an IP name is specified as host.domain (e.g., uscms.fnal.gov)
  - 131.225 is registered as the fnal.gov domain
  - 52.112 is the uscms host address within the fnal.gov domain
    - 52 (xxx.yyy.52.zzz) is the Fermilab CMS sub-domain address
    - 112 (xxx.yyy.ww.112) is the host address

URLs

- File locations on a server are specified relative to a root (/) location on the server
- The root location is not necessarily at the root of the file system, but rather is specified by the server for each protocol (and for each client)
- The general form of a URL (uniform resource locator) is:
  - protocol://username:password@host.domain:port/directory/.../file.ext
  - For example, http://uscms.fnal.gov/ specifies the root of the WWW directory structure on host uscms.fnal.gov
    - » no username or password
    - » default http port (80)
    - » default file (or action) in the root directory, because no file is specified

File Sharing

Files may be shared among computers by mounting a remote computer's file structure on a local computer

For example, portions of the US CMS server's file system may be mounted via AppleShare or Microsoft Networking on another computer

- Macintosh: Chooser...AppleShare...CMS/MAIN ETHERNET...US CMS Server
- Windows: Start...Find...Computer...US_CMS
- A symbolic link file (alias or shortcut) may be created on the remote desktop for ease of subsequent access
- Access to files on the US CMS server is password controlled

The US CMS Server

The US CMS server (IP name uscms.fnal.gov) is configured as an ftp, http (WWW), pop3, smtp, and FileMaker Pro server

In addition, files on the US CMS server may be shared via AppleShare and Microsoft Networking

We will concentrate on serving files via the World-wide Web and via ftp; and on AppleShare and Microsoft file sharing

The US CMS File System

The file system on the US CMS server is configured with two independent WWW roots (US CMS and Project Office), and a multi-rooted ftp service

The directory structure on the US CMS server is:

- / (root directory on the hard disk)
  - Servers
    - WWW (US CMS WWW root)
      - pub (anonymous ftp - read only)
      - uscms (all US CMS WWW files except the home page)
    - WWWadmin (Project Office WWW root)
      - project_office (all PO WWW files except the home page)
    - incoming (semi-anonymous ftp [username uscms] world read/write)
    - ftp (anonymous ftp root linked to other ftp roots - read only)

The Project Office File System

The Project Office file system (WWWadmin root) is configured with a number of different levels of read/write access, most of which are password protected

- / (root directory on the hard disk)
  - Servers
    - WWW (US CMS WWW root)
    - WWWadmin (Project Office WWW root)
      - project_office (all PO WWW files except the home page)
        - cs_proc (cost/schedule change procedures)
        - l1m (project office documents)
          - dropbox (anonymous ftp - write only)
        - l2m (Microsoft Project files, MOUs, ...)
        - subsystem (SOWs, ...)
    - incoming (anonymous ftp [username uscms] world read/write)
    - ftp (anonymous ftp root linked to other ftp roots - read only)

Serving Files via ftp

Ftp file servers have been around for many years, and ftp remains the fastest and most reliable method to transfer files between computers

Files are transmitted in either text (ASCII) or binary (raw) formats

Advantages

- fast
- reliable
- most commonly available client
- preferred method for transferring large files

Disadvantages

- text oriented interface
- file oriented transfer
- non-encrypted passwords

The ftp Client

- A typical ftp installation provides for both user (username and password) and anonymous ftp sessions
  - Anonymous ftp is an ftp session with username “anonymous”, and a user’s e-mail address as the password
  - Usually, any or no password is accepted
- The default action for an ftp client is to list the files in a directory on the host. The client may either “get” or “put” a file (in text or raw format) assuming sufficient privilege, or change to another directory on the host

ftp on the US CMS Server

- The bulk of the files accessible via ftp on the US CMS server are within the /pub ftp subdirectory
  - /pub is also within the http (WWW) root, so these files are also available via the World-wide Web
- In the Project Office file system, only the dropbox is accessible via ftp
  - Special anonymous-like ftp service
  - username of “admin” and any or no password
  - write-only permission

Serving Files via http

- Files are served on the World-wide Web via hypertext transport protocol, or http
- Files must reside within the defined http root on the server to be accessible to web clients
- A client WWW browser (e.g., Navigator or Explorer) sends a URL request to a web server
- The server returns the requested file, or performs a default action if the URL specifies a directory, rather than a file
  - The default action in a directory could be to return a default file, or to return a directory listing
- If the requested file contains hypertext markup language (html), the WWW client will display the content formatted according to html specifications

The http Client

- An http client, or WWW browser, such as Netscape Navigator or Internet Explorer typically receives a requested html file as text, and formats the display according to html specifications and according to the user's own preferences
  - The server has no control over the client's software, software version, and the user's settings, and thus there is no guarantee that a web page will display as intended on all clients

http on the US CMS Server

On the US CMS server, the default action if a URL specifies a directory is:

- return a directory listing if and only if the directory contains a file named ".message", and display the content (if any) of the .message file as text
- or -
- return an html file named "default.html", if such a file exists
- or -
- return an error message

The option to upload files via http is disabled on the US CMS server; ftp upload is used instead

Writing html Files

- A basic html document can be written using a simple text editing program; a template is available in the WWWadmin/project_office folder on the US CMS server
  - http://uscms.fnal.gov:8001/project_office/template.html
- White space (spaces, tabs, returns, ...) is treated as a single space character within html files
- Formatting derives entirely from directives enclosed between pairs of html tags
  - <tag> ... </tag>
- Links to other files are specified by including the URL of the referenced file within the parent html file using the <A HREF="URL"> ... </A> tag

```
<HTML>

<HEAD>
<!-- Put page title here -->
<TITLE>Page Title</TITLE>
</HEAD>

<!-- Set body background to white. -->
<BODY BGCOLOR="#FFFFFF">

<!-- US CMS Project Office Banner -->
<center><img src="http://uscms.fnal.gov:8001/project_office/banner_uspo.gif"></center><br>

<!-- Horizontal Rule -->
<HR SIZE="2">

<!-- Level 2 heading -->
<H2>Level 2 Heading</H2>

<!-- Text with line break -->
Descriptive text can be added as desired. <br>

<p>Text can also be enclosed in paragraph tags.</p>

<!-- Level 3 heading -->
<H3>Level 3 Heading</H3>

<!-- Unordered List -->
<ul>
<li>List item
<li>List item with <A HREF="http://uscms.fnal.gov/">link to http://uscms.fnal.gov/</A>
<li>List item with <A HREF="../">link to parent directory</A>
</ul>

</BODY>
</HTML>
```

Posting Files on the Web

- Many programs (Microsoft Word, Excel, ...) contain options to “save as” an html file
- To be useful for display on the web, the resulting .html file must be linked (using the <A...> </A> tags) to an existing html file
  - Often, the resulting html file needs to be tweaked to achieve the desired display on the web
- Given a linked subdirectory on a web server which returns a directory index as the default action, additional files may be posted on the web simply by adding (copying) them to the existing subdirectory
  - When a user clicks on a file name in a directory listing, the default server action is to return that file to the user’s web browser
- File names should contain no special characters

Serving Different File Types

- Any file of file type may be included in an indexed directory listing, or linked with a URL in an html file
- The result when a user clicks on a link or a file name depends on the capabilities and settings of both the server and client software
- The action depends on the file type, or extension (html, gif, jpg, ps, eps, pdf, doc, xls, mpp, ...) of the file being served
  - All modern web browsers (clients) will interpret and translate the html directives in an html file, and will display an image from a gif file
  - Most web browsers are capable of displaying text and graphics in a pdf file
  - Almost no web browsers know what to do with a mpp file

Summary and Conclusions

The US CMS ftp and web servers are available for the US CMS Collaboration and for the US CMS Project Office

Files may be posted on the server through drag and drop copying or in situ editing by mounting the US CMS files via AppleShare or Microsoft Networking on the local desktop

Keep in mind that the clients will often not have the latest web browser and application software versions available

Above all, experiment, learn by trying and doing, and ...

... have fun doing it!